A white ceramic cup filled with a light brown coffee, sitting on a matching white saucer. The cup has a simple, modern design with a curved handle. The background is a plain, light color.

EXPLORATORY DATA ANALYSIS AND INSIGHTS GATHERING (With special focus on gender diversity and pay gap)

Using SQL

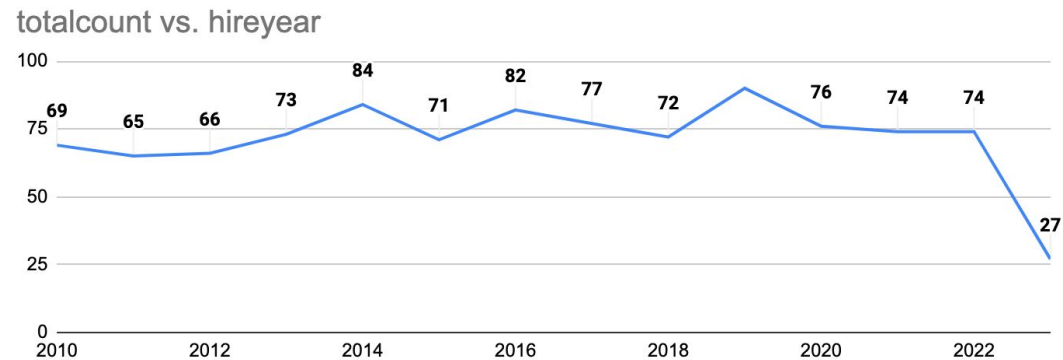
Coffee shops employees data

Bala Mira G

-----Insights-----

-----Hiring across the years -----

```
SELECT  
  COUNT(*) AS totalcount,  
  YEAR(hire_date) AS hireyear  
FROM  
  employees  
GROUP BY hireyear  
ORDER BY hireyear;
```



Insights gathered:

There has been a same trend in hiring observed over the years except for a very low dip in 2023.

-----Analyzing 3 coffee shops sales and their employees' data-----

-----Gender Bifurcation and %-----

```
SELECT
  gender,
  COUNT(*) AS count,
  ROUND((COUNT(*) / (SELECT
    COUNT(*)
    FROM
      employees)) * 100,
    0) AS percentage
FROM
  employees
GROUP BY
  gender order by percentage desc;
;
```

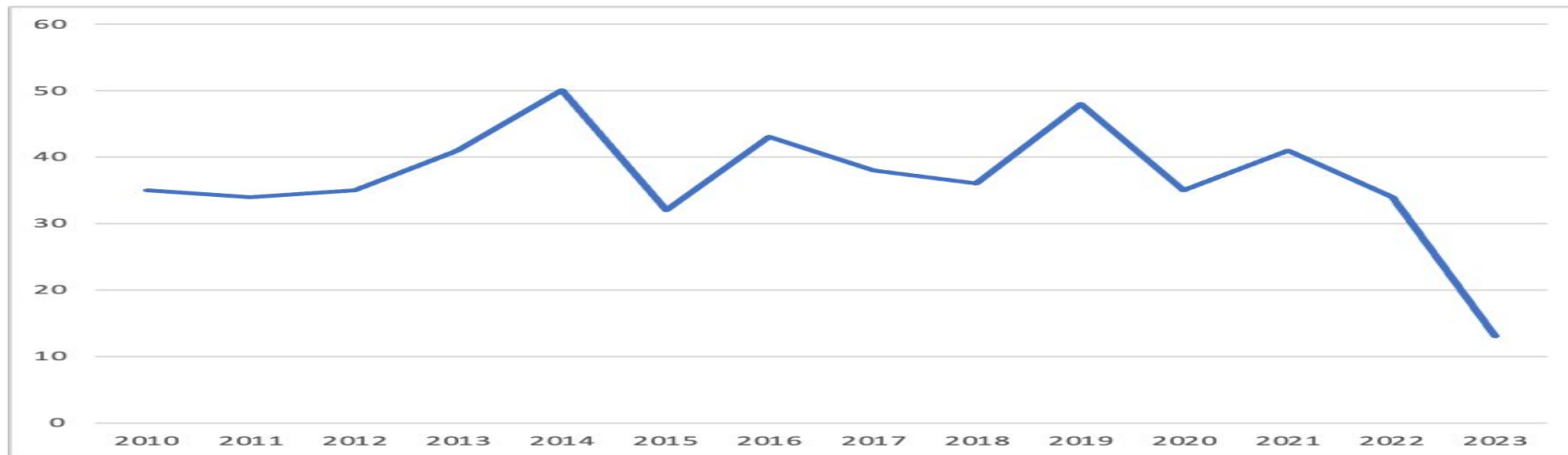
```
SELECT
  COUNT(*) AS totalcount, gender
FROM
  employees
GROUP BY gender
ORDER BY totalcount DESC;
```

Insights gathered:

Males form 49% and Females form 52% of the total employees in all the three coffee shops which indicates a good improvement when it comes to gender diversity.

----- Female Employee count yearly trend -----

```
SELECT  
  COUNT(*) as femalecount, gender, YEAR(hire_date) AS hireyear  
FROM  
  employees  
WHERE  
  gender = 'F'  
GROUP BY hireyear  
ORDER BY hireyear ASC;
```

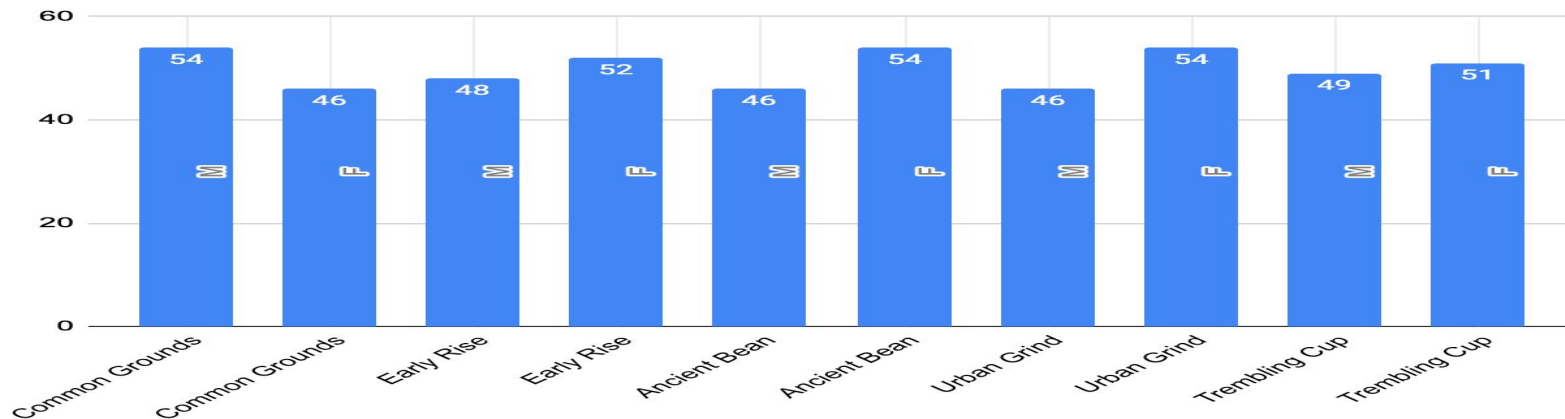


Insights gathered:

Female employee count has been the lowest in 2023 while other years have exhibited the same trend.

----- Female Employee count company wise trend-----

```
with cte as (SELECT
  `coffeeshop_id`,
  gender,
  round((COUNT(*) * 100.0 / SUM(COUNT(*)) OVER (PARTITION BY `coffeeshop_id`)),0) AS
  genderpercentage
FROM
  employees
GROUP BY
  `coffeeshop_id`, gender)
select s.`coffeeshop_name`, c.gender,c.genderpercentage from cte c join shops s on
c.`coffeeshop_id`=s.`coffeeshop_id`;
```

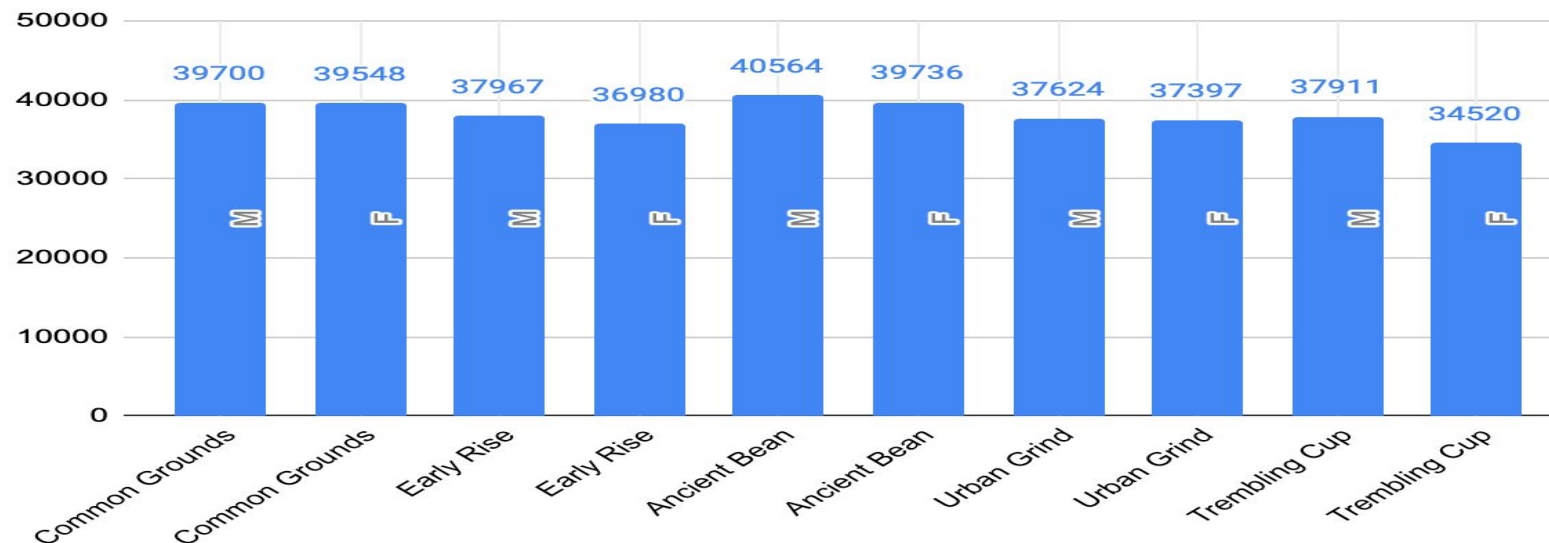


Insights gathered:

Female employee % has been higher than that of male in all the coffeeshops except Common Grounds

----- Salary Analysis for female employees -----

```
with cte as (select
  round(AVG(salary),0) as avgsal, gender, `coffeeshop_id`
FROM
  employees
GROUP BY gender, `coffeeshop_id`)
select s.`coffeeshop_name`,c.gender,c.avgsal from cte c join shops s on
c.`coffeeshop_id`=s.`coffeeshop_id`;
```

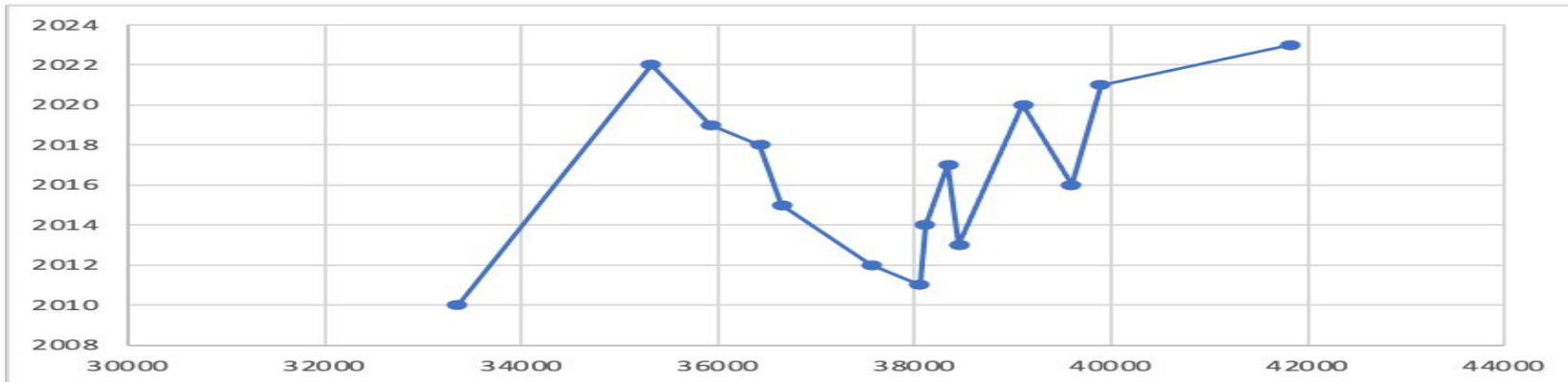


Insights gathered:

The avg salary offered to female employees in all the coffeeshops is lesser than that of their male counterparts.

----- Salary Analysis for Female employees -----

```
with cte as (SELECT
  ROUND(AVG(salary), 0) AS avgsal,gender,hireyear
FROM
  (SELECT
    e.salary, year(e.`hire_date`) as hireyear ,gender
  FROM
    employees e
  JOIN shops s ON e.`coffeeshop_id` = s.`coffeeshop_id`) a
GROUP BY gender,hireyear
ORDER BY avgsal DESC)
select * from cte where gender="F";
```



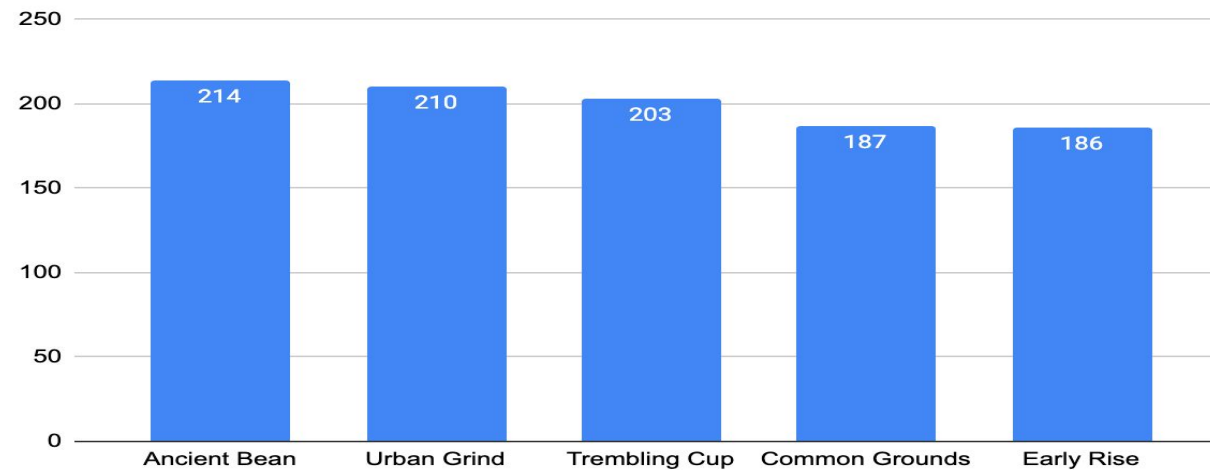
Insights gathered:

Average salary offered to female employees has hit the lowest in 2011. Though 2023 had the lowest female hires the average salary is still high.

----- Employee count per coffee shop -----

```
with cte as (select s.`coffeeshop_name` from employees e join shops s on  
e.`coffeeshop_id`=s.`coffeeshop_id`)  
select `coffeeshop_name`,count(*) as totalcount from cte group by `coffeeshop_name` order by  
totalcount desc;
```

totalcount vs. coffeeshop_name

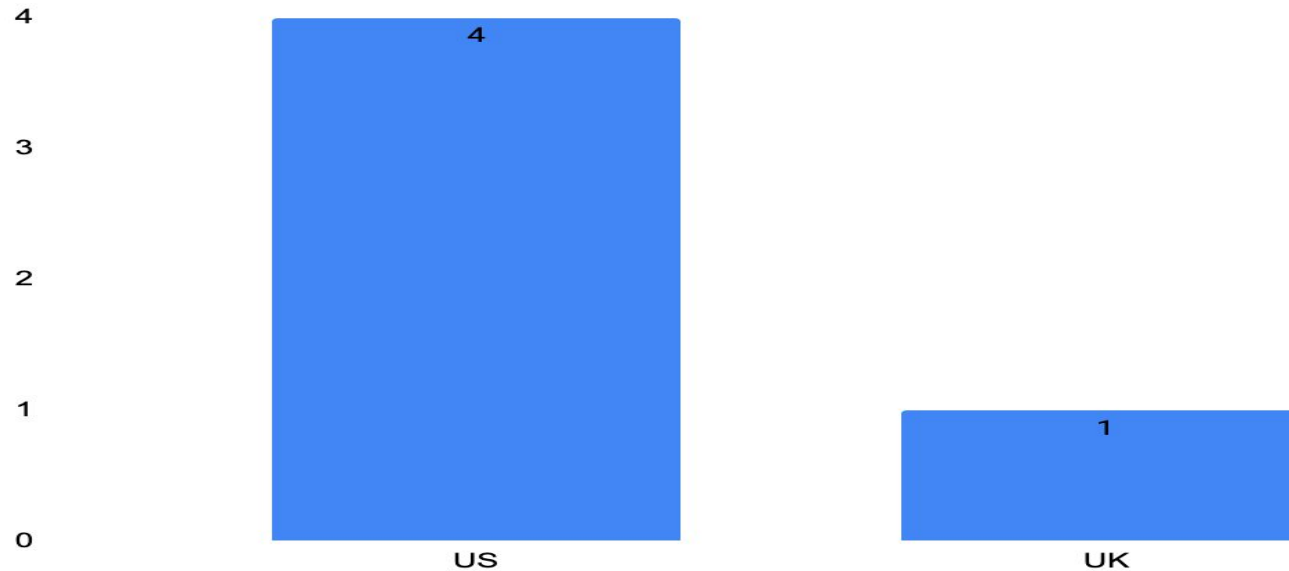


Insights gathered:

Ancient Bean has been the largest employer whereas Early Rise has been the lowest.

----- Shop count per country -----

```
with cte as (select l.country ,l.city from shops ss join locations l on ss.`city_id`=l.`city_id`)  
select count(*) as totalshops,country from cte group by country order by totalshops desc;
```

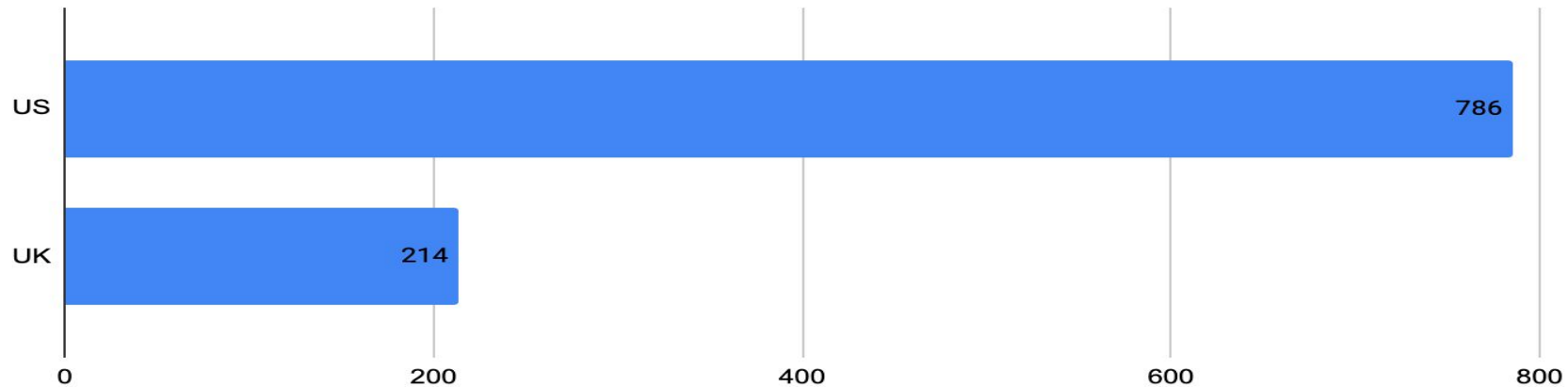


Insights gathered:

US has higher no of coffee shops than the UK.

----- Employee count per country -----

```
select count(*) as totalemployeecount,country from (SELECT  
  l.country  
FROM  
  employees e  
JOIN shops s ON e.`coffeeshop_id` = s.`coffeeshop_id`  
JOIN locations l ON l.`city_id` = s.`city_id`) a group by country; se
```



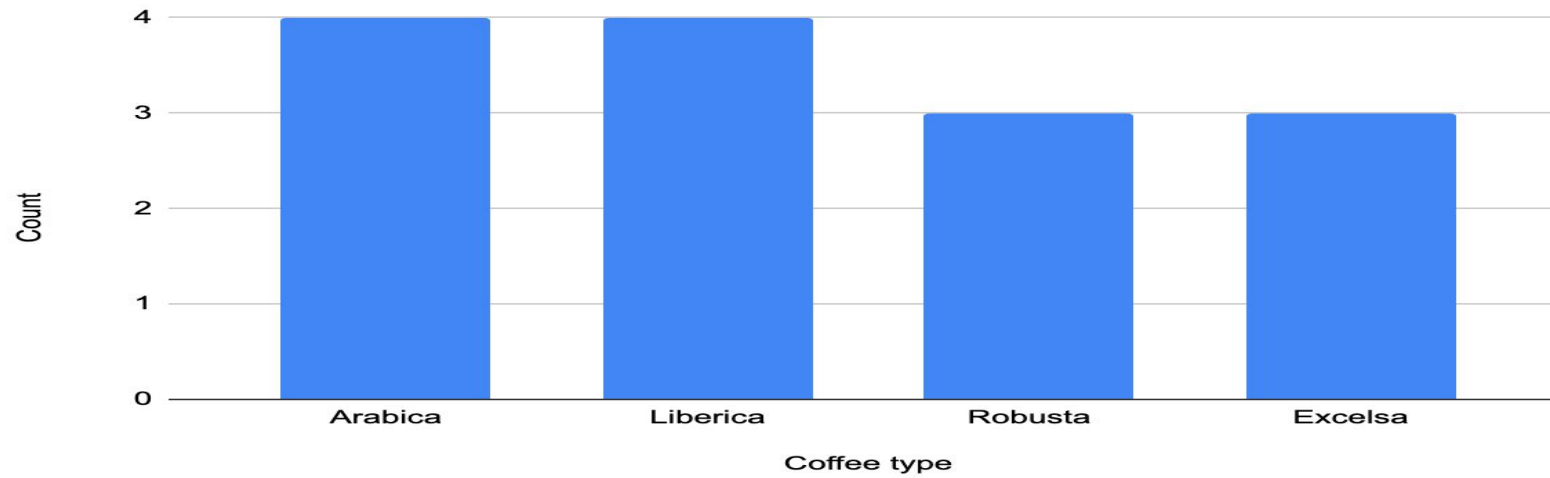
Insights gathered:

US has higher no of employees than the UK.

----- Most preferred coffee type by the supplier -----

```
SELECT
  COUNT(*) AS beancount, `coffee_type`
FROM
  suppliers
GROUP BY `coffee_type`
ORDER BY beancount DESC;
```

Count vs. Coffee type



----- Insights gathered and recommendations -----

Insights and recommendations:

- 2023 has seen a dip in hiring both in terms of male and female hiring
- Data gathered from the coffee shops indicate that female hiring % has been more than the male.
- On the contrary, avg salary offered to the female employees are lower than that of male.
- Avg salary for female employees hit the highest in 2023 though the hiring number is low.
- Steps can be taken by the coffee shops to improve the gender pay gap thereby ensuring pay parity remains equal between the genders